

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平10-336354

(43) 公開日 平成10年(1998)12月18日

(51) Int.Cl.<sup>6</sup>

識別記号

F I

H 0 4 M 17/00

H 0 4 M 17/00

Z

G 0 6 F 17/28

G 1 0 L 5/00

G 1 0 L 5/00

H 0 4 M 11/06

H 0 4 M 11/06

G 0 6 F 15/18

5 6 0 G

// G 0 6 F 15/18

5 6 0

15/38

V

審査請求 未請求 請求項の数 2 O L (全 10 頁)

(21) 出願番号

特願平9-146123

(22) 出願日

平成9年(1997)6月4日

(71) 出願人 000006105

株式会社明電舎

東京都品川区大崎2丁目1番17号

(72) 発明者 高橋 和彦

東京都品川区大崎2丁目1番17号 株式会社明電舎内

(74) 代理人 弁理士 志賀 富士弥 (外1名)

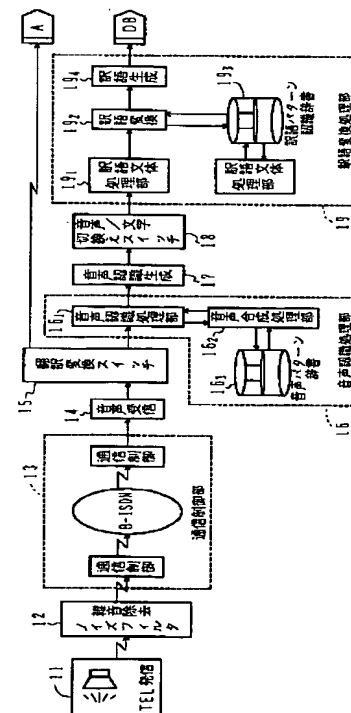
(54) 【発明の名称】 マルチメディア公衆電話システム

(57) 【要約】

【課題】 マルチメディア公衆電話システムは、映像も送受信できるが、異なる言語による直接の会話ができない。

【解決手段】 通話者の音声を通信システムを介して音声信号受信手段11～14で受信し、翻訳変換スイッチ15で翻訳を選択したとき、音声認識装置16により受信した音声信号を認識し、この認識音声を生声認識生成装置17で文字データに変換し、音声／文字切換スイッチ18で音声／文字の出力モードを変え、訳語変換処理部19で翻訳し、翻訳結果を音声又は文字でディスプレイ上に表示する。逆に、通話者の音声又は入力した文字を音声又は文字認識し、これを翻訳して音声合成し、通信システムを介して通話相手に送信する手段も含む。

マルチメディア公衆電話の処理ブロック図(その1)



## 【特許請求の範囲】

【請求項1】 音声と映像の入出力手段と通信手段を有して互いに音声及び映像を使った通話を可能にしたマルチメディア公衆電話システムにおいて、通話者の音声を通信制御部を介して受信する音声信号受信手段と、前記音声信号をそのまま音声で出力又は翻訳して出力するかを選択する翻訳変換スイッチと、前記スイッチが翻訳を選択したとき、受信した音声信号を認識する音声認識装置と、前記音声認識装置が認識した音声を通話者が指定した言語に翻訳する訳語変換処理装置と、前記翻訳した訳語を音声又は文字でディスプレイ上に文字で表示する変換装置とを備えたことを特徴とするマルチメディア公衆電話システム。

【請求項2】 音声と映像の入出力手段と通信手段を有して互いに音声及び映像を使った通話を可能にしたマルチメディア公衆電話システムにおいて、通話者の音声又は入力文字をそのまま相手に送信するか、又は翻訳して音声又は文字で送信するかを選択する音声／文字変換スイッチと、前記翻訳する訳語を選択する訳語選択部と、前記スイッチが文字送信を選択したとき、通話者が入力した文字列を文字認識により文体として認識生成する文体認識装置と、前記スイッチが音声送信を選択したとき、通話者が入力した音声を認識する音声認識装置と、前記認識された文体又は音声を前記訳語選択部で選択した言語に翻訳する訳語変換処理装置と、前記翻訳した訳語を音声合成により音声波形に変換する音声合成装置と、前記音声合成装置が合成した音声波形を音声信号に変換して通信手段を介して通話相手に送信する音声変換装置とを備えたことを特徴とするマルチメディア公衆電話システム。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】 本発明は、公衆電話システムにおいて、送信側と受信側で異なる言語になる音声又は文字を翻訳して送受信できるようにしたマルチメディア公衆電話システムに関する。

## 【0002】

【従来の技術】 現在のマルチメディア公衆電話システムは、図16に示すように、CCDカメラ1による映像入力とスクリーン2による映像出力になる映像入出力装置により、相手の顔を見ながら同時に会話ができること、FAX機能を持つこと、設置場所の周辺マップや情報案内機能を持つこと、メモ欄機能を持つこと等の機能が加えられ、さらに天気予報が聞けること、各イベントの情報を取り寄せること等の各種の機能がある。この公衆電

話の操作入力は、プッシュボタン3とポインティングデバイス（ペン入力やタッチスクリーン方式）4が利用される。

## 【0003】

【発明が解決しようとする課題】 従来のマルチメディア公衆電話システムを利用して海外諸国とコミュニケーションを行うとき、国際電話と同様に言語の意味が解らなかつたり、聞き逃し、聞き間違い、そして翻訳サービス等へ依頼したりと、コミュニケーション間のギャップがあつてスムーズにいかないことがあり、以下のような問題がある。

【0004】（1）送受信者が共に英語や日本語で話す場合のように、同じ言葉でしかコミュニケーションが取れない。

【0005】（2）送受信者が異なる言葉を使う場合には言語間の通訳を必要とする。

【0006】（3）翻訳サービスを通訳としてコミュニケーションを行うと、プライバシーの侵害などで会話がやりにくいし、翻訳のための時間がかかる。

【0007】（4）言語間のギャップなどで会話途中にあいまいな発声や言語の意味が解らない場合や聞き逃す場合があり、円滑で確実なコミュニケーションが難しい。

【0008】 以上のように、従来のシステムは種々のマルチメディア機能があつても送受信者が異なる言葉を使う場合には利用できる範囲が限られている。

【0009】 本発明の目的は、送受信者が異なる言葉を使う場合にも通訳を介することなく円滑で確実なコミュニケーションができるマルチメディア公衆電話システムを提供することにある。

## 【0010】

【課題を解決するための手段】 本発明は、上記課題を解決するため、従来の機能に異なる言語の音声又は文字への自動翻訳機能を追加し、翻訳モードの選択で相手の音声又は入力文字を認識してそれを翻訳し、翻訳結果を音声合成して音声出力すること、又は字幕に変換処理してディスプレイ上に表示するようにしたもので、以下の構成を特徴とする。

【0011】（第1の発明） 音声と映像の入出力手段と通信手段を有して互いに音声及び映像を使った通話を可能にしたマルチメディア公衆電話システムにおいて、通話者の音声を通信制御部を介して受信する音声信号受信手段と、前記音声信号をそのまま音声で出力又は翻訳して出力するかを選択する翻訳変換スイッチと、前記スイッチが翻訳を選択したとき、受信した音声信号を認識する音声認識装置と、前記音声認識装置が認識した音声を通話者が指定した言語に翻訳する訳語変換処理装置と、前記翻訳した訳語を音声又は文字でディスプレイ上に文字で表示する変換装置とを備えたことを特徴とする。

【0012】（第2の発明） 音声と映像の入出力手段と

## 3

通信手段を有して互いに音声及び映像を使った通話を可能にしたマルチメディア公衆電話システムにおいて、通話者の音声又は入力文字をそのまま相手に送信するか、又は翻訳して音声又は文字で送信するかを選択する音声／文字変換スイッチと、前記翻訳する訳語を選択する訳語選択部と、前記スイッチが文字送信を選択したとき、通話者が入力した文字列を文字認識により文体として認識生成する文体認識装置と、前記スイッチが音声送信を選択したとき、通話者が入力した音声を認識する音声認識装置と、前記認識された文体又は音声を前記訳語選択部で選択した言語に翻訳する訳語変換処理装置と、前記翻訳した訳語を音声合成により音声波形に変換する音声合成装置と、前記音声合成装置が合成した音声波形を音声信号に変換して通信手段を介して通話相手に送信する音声変換装置とを備えたことを特徴とする。

## 【0013】

【発明の実施の形態】図1～図4は、本発明の実施形態を示すマルチメディア公衆電話システムのブロック図であり、図5～図9に各部の処理過程を同じ符号で対応付けて示す。以下、各部の構成と処理を詳細に説明する。

【0014】(1)相手の音声を翻訳して受信する場合(図1及び図2)。

【0015】音声入力装置11は、公衆電話に搭載される受話器のマイクロフォンから音声を入力するためのものであり、通話者が受話器から音声を入力した信号を発生する。雑音除去フィルタ12は、会話の中に紛れて入ってくる雑音をノイズフィルタで除去する。

【0016】通信制御部13は、例えば、ディジタル回線で高速かつ広帯域性がある2B+DのB-ISDN及びネットワークを制御するATM交換システムによって通信制御を行い、音声入力装置11からの音声信号を音声受信装置14に伝送する。

【0017】翻訳変換スイッチ15は、受信した音声(アナログ信号)をそのまま音声として受信するか、受信者が理解できる言語に翻訳して文字又は音声で受信するかを受信者が選択する。

【0018】音声理解装置16は、スイッチ15の選択操作が翻訳受信にされた場合に受信音声を認識及び理解する。この音声理解装置16は、受信した音声を認識及び理解するため、不特定話者に対応可能な技術として音声分析とファジィ制御によるあいまいな言語を除去する特徴抽出後に単語・構文・意味解析による音声認識を行うシステムに構成され、音声認識処理部16<sub>1</sub>と音声合成処理部16<sub>2</sub>及び音声パターン辞書16<sub>3</sub>を備える。

【0019】この装置16では、図6及び図7の処理フローで示すように、入力された音声信号をA/D変換処理した後、音声分析により音声入力信号を周波数スペクトルによってファジィ制御であいまいな音声を除去する。そして、特徴抽出によって音声分析結果から音声信号の特徴パラメータを時系列に抽出・変換し、セグメン

## 4

テーション処理により音声単位へのセグメント化を行う。

【0020】特徴抽出には、「えーと」や「あー」、「うーん」のように、冗長的な単語をファジィ推論により抽出して除去し、人と人が会話をする上で本当に必要な単語のみを抽出する。さらに、日常的に必要な単語は前もって登録しておき、後に必要となった単語を学習機能によってしだいに認識及び追加して行く。

【0021】次いで、音声認識によりニューラルネットワーク技術を使って音声標準パターンとの比較で単語認識を行うことで音素系列を得、単語照合と単語認識により音素系列について知識ベースに持つ単語標準パターンとの照合で単語を認識する。この認識に標準パターンが存在しなければ学習処理(認識・登録)を行って単語知識を知識ベースに追加する。

【0022】次いで、認識した単語について、構文照合と構文認識により知識ベースの構文パターンとの照合で構文的に誤りがあるか否かを解析し、誤りがあれば再検証することで構文認識を行う。さらに、意味解析と意味認識により認識された単語構文について意味的に妥当であるかを調べ、妥当な結果が得られるまで構文解析処理と意味解析処理を繰り返す。これら解析には文字に変換可能か否かも含め、文字変換処理により変換可能である文字を漢字やカナに変換する。

【0023】図1に戻って、音声認識生成装置17は、音声理解装置16によって受信音声の内容が認識された音声データをテキスト形式の文字データに変換する。音声／文字切換えスイッチ18は、文字データを翻訳した内容を文字で表示するか／音声で出力するかを受信者が選択する。

【0024】訳語変換処理部19は、音声認識された受信内容をスイッチ15で設定する言語に翻訳する。この翻訳は、訳語文体処理部19<sub>1</sub>によって受信内容について意味辞書を使い学習機能を持たせて意味解析と意味認識を行って文字列を認識する文体処理を行い、訳語変換部19<sub>2</sub>が訳語パターン認識辞書19<sub>3</sub>を参照した訳語パターンにしたがって翻訳し、訳語生成部19<sub>4</sub>で翻訳した訳語をつなぎ、翻訳文を生成する。

【0025】変換処理部20A、20Bは、翻訳文を文字コードに変換又は音声合成による音声データに変換し、信号変換部21A、21Bはこれら変換データを実際の文字又は音声に変換する。表示部22Aは文字を公衆電話のスクリーン2に表示し、スピーカ出力部22Bは音声信号を増幅して電話の受話器スピーカに出力する。

【0026】したがって、図1及び図2の構成により、受信した音声を音声認識装置により音声認識を行い、これを受信者が希望する言語に翻訳して文字又は音声として受信することにより、発信者と異なる言語による会話でも受信者は自分の言語を使った会話ができる。

【0027】また、電話でのコミュニケーション中、スクリーン表示により聞き漏らしなどのリスクを回避することができる。

【0028】(2) 自分の音声又は入力文字を翻訳して相手に送信する場合(図3及び図4) 音声/文字切換スイッチ23は、会話の受信者が、送信者に返答するために、音声又は文字で書き表した内容を翻訳してコミュニケーションを行うときにその操作で音声/文字切換えのボタンを操作する。

【0029】文字入力装置24Aは、文字を選択したときにメモ帳(テキスト形式)を画面表示し、話の内容をポインティングデバイス(電子ペン)で入力する。音声入力装置24B 音声を選択したときに受話器から入力する音声信号を得る。訳語選択部25は、入力された文字又は音声をどの言語に翻訳するかを選択する。

【0030】文体認識装置26は、手書き入力した文字列を解析して文章として認識するもので、入力が文字の場合にその文体を認識して生成する。この認識/生成には、文体形式分析26<sub>2</sub>による入力文体形式の分析を介した文字パターンとその認識辞書26<sub>1</sub>のパターンと比較することで文字認識部26<sub>3</sub>が文字認識を行い、この文字認識結果から文体認識生成部26<sub>4</sub>が文体として生成する。

【0031】音声認識処理部27及び音声認識生成部28は、図1の16及び17をそのまま流用し、入力音声を認識する。同様に、訳語変換処理部29は、図1の19をそのまま流用し、入力音声信号又は文字を翻訳する。

【0032】音声合成処理部30は、27と同様の処理で、スイッチ23で選択した言語に翻訳した文字又は音声信号を音声合成する。音声変換部31は、合成した音声信号を実際の音声信号に変換する。

【0033】通信制御部31は、図1の13と同様に公衆回線を使って相手に送信する。この送信音声は、相手の音声受信部33を経て受話器から翻訳した音声として出力される。

【0034】したがって、図3及び図4の構成により、相手が話す言語に翻訳して相手に音声で送信することができる。

【0035】なお、以上までのブロック構成において、音声合成装置は、例えば規則合成方式により音声合成を行い、漢字カナ混じり文に対して日本語辞書を参照した構文解析、意味解析等により読み・単語・文節境界等を解析し、この解析結果から音声パターン辞書を参照して音韻系列のアクセントとイントネーション及び音韻の継続時間の各パラメータを決定し、これらを音響管モデルのパラメータとして音声合成を行うことができる。

【0036】また、訳語翻訳処理部19、29は、トランスファー方式、ダイレクト方式、ピボット方式、さらにはこれらの方式の特徴を組み合わせた方式で実現され

る。

【0037】トランスファー方式は、言語別の中間表現を持ち、中間表現での変換が行われる。例えば、英語→日本語、仏語→日本語などを翻訳する言語に中間表現がそれぞれ生成される。また、解析手法としては日本語には格文法、英語には拡張遷移文法がよく利用される。

【0038】図10は、トランスファー方式による翻訳システム構成を例文と共に示す。元の言語Aとしての日本語の文「明日は、晴れでしょう」を目的の言語Bとしての英語の文に変換するのに、元の文を単語辞書C及び文法辞書Dを使った日本語処理方法による構文解析と意味解析を行う構文・意味解析Eを行うことで文節「明日は」と「でしょう」と「晴れ」の組み合わせになる中間表現Fを得る。

【0039】この中間表現Fを「tomorrow」と「It will」と「fine」の組み合わせになる英語の中間表現Gに変換する。この中間表現Gから辞書C、Dを使った英語の表層文生成Hにより目的の言語B「It will fine tomorrow」を生成する。

【0040】図11は、トランスファー方式とダイレクト方式及びピボット方式の3つの方式を組み合わせた変換方式を示す。トランスファー方式は、前記のように元の言語Aと変換目的の言語Bにそれぞれ特徴的な中間表現F、Gを生成した変換を行う。

【0041】ダイレクト方式は、元の言語の単語を目的の言語の単語に置き換え、その単語の順序を並べ換えて変換し、ピボット方式は、元の言語の種類に依存しない共通の中間表現Iに変換し、その中間表現から目的とする言語を生成するが、これら両方式は、トランスファー方式に比べて変換精度が劣る。

【0042】なお、上記の3つの変換方式の特徴を合わせ、より複雑な翻訳処理を可能にする方式もある。

【0043】図12は、画面レイアウト例である。この画面はマルチメディア機能の他、映像部41、音声を字幕に切り替えた受信内容テキスト画面42、相手に文字を入力して音声で内容を送信するための送信内容テキスト画面43、翻訳した言語の選択画面44及び入力した内容を送信する時、音声を男性の声が女性の声を選択する音声モード45の5つからなる。

【0044】なお、本発明は、図1の構成又は図2の構成の両方の装置を搭載するマルチメディア公衆電話システムとする他に、図1又は図2の一方の装置のみを単独に搭載するマルチメディア公衆電話システムとすることもできる。

【0045】図13は、相手の音声を字幕表示で受信する機能と、文字入力を相手に音声で送信する機能の両方を持つマルチメディア公衆電話システムのインタフェース構成例を示す。

【0046】同図の音声入力部は、図14に処理手順で

10

20

30

40

50

示すように、マイクロフォンから音声を入力し（S1）、ノイズフィルタで音声に交じっている雑音を除去し（S2）、音声入力ボードによりクリアな音声をデジタル信号に変換し（S3）、音声入力ドライバの制御の基にニューラルネットワークによる学習機能を有して音声認識し（S4）、ファジィ推論部の推論により認識された音声信号に冗長な信号を除去し（S5）、知識ベースと知識ベースファイルと自動プログラミングエディタにより認識された音声信号が知識ベースファイルに無い場合に学習処理を行い記憶し（S6）、ウィンドウコントロールにより最終的に音声信号処理が完了・認識された結果をウィンドウのエディタアプリケーション画面に表示する（S7）。

【0047】図13の文字入力部は、図15に処理手順で示すように、テキストエディタとペンにより文字入力し（S11）、入力された文字のパターン処理で認識するため、文字パターン認識部と学習部と専用知識ベース及び汎用知識ベースエディタにより、入力した文字のパターンやデータベース（db）に記憶されていない単語を学習して記憶し、そして文字のパターンが判別されることで認識する（S12）。次いで、認識された文字（デジタル）を音声に変換する（アナログ）ため、音声合成システムにより男性の声又は女性の声で音声合成し（S13）、音声合成処理を完了した結果を音声出力部（マイクロホン）で出力する（S14）。

#### 【0048】

【発明の効果】以上のとおり、本発明によれば、従来のマルチメディア公衆電話機能に異なる言語の音声又は文字への自動翻訳機能を追加し、翻訳モードの選択で相手の音声又は入力文字を認識してそれを翻訳し、翻訳結果を音声合成して音声出力すること、又は字幕に変換処理してディスプレイ上に表示するようにしたため、以下の効果がある。

【0049】（1）通話者と相手が互いに異なる言語を使って直接に会話できる。

【0050】（2）会話中、重要と思われる点を字幕で表示させる事ができるため、聞き違いなどのリスクを回避する事が可能になる。

【0051】（3）コンピュータが翻訳処理を行うた

め、翻訳サービスを利用する場合に比べてプライバシーが守られる。

【0052】（4）コミュニケーション上のギャップを回避することが可能となる。

#### 【図面の簡単な説明】

【図1】本発明の実施形態を示すマルチメディア公衆電話システムのブロック構成図（その1）。

【図2】本発明の実施形態を示すマルチメディア公衆電話システムのブロック構成図（その2）。

10 【図3】本発明の実施形態を示すマルチメディア公衆電話システムのブロック構成図（その3）。

【図4】本発明の実施形態を示すマルチメディア公衆電話システムのブロック構成図（その4）。

【図5】実施形態における処理フロー（その1）。

【図6】実施形態における処理フロー（その2）。

【図7】実施形態における処理フロー（その3）。

【図8】実施形態における処理フロー（その4）。

【図9】実施形態における処理フロー（その5）。

【図10】翻訳システムの構成図。

20 【図11】他の翻訳システムの構成図。

【図12】実施形態における画面レイアウト例。

【図13】実施形態におけるインタフェース構成例。

【図14】図13における音声入力部の処理手順。

【図15】図13における文字入力部の処理手順。

【図16】現在のマルチメディア公衆電話の図。

#### 【符号の説明】

11…音声入力装置

12…雑音除去ノイズフィルタ

13、32…通信制御部

30 14…音声受信装置

15…翻訳変換スイッチ

16、27…音声認識処理部

17、28…音声認識生成装置

18…音声／文字切換スイッチ

19、29…訳語変換処理部

23…音声／文字切換スイッチ

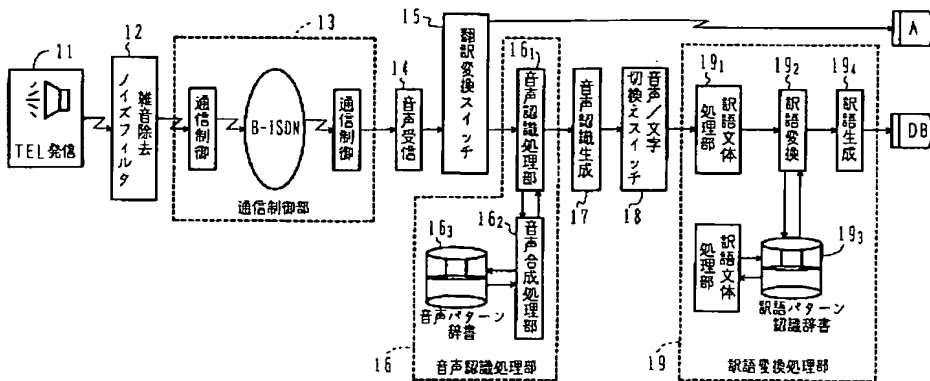
25…訳語選択部

26…文体認識装置

30…音声合成処理部

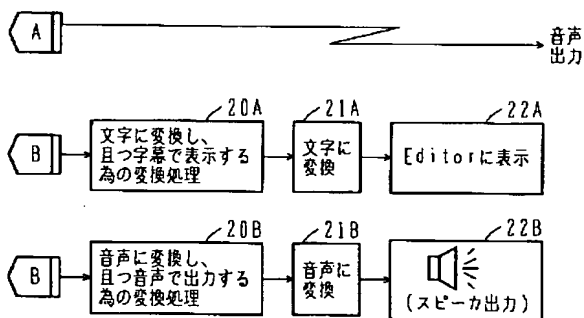
【図1】

マルチメディア公衆電話の処理ブロック図(その1)



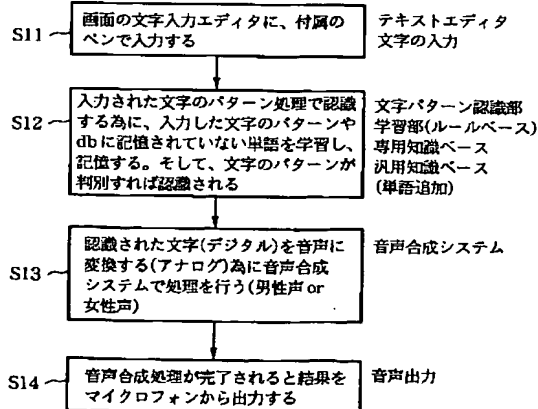
【図2】

マルチメディア公衆電話の処理ブロック図(その2)



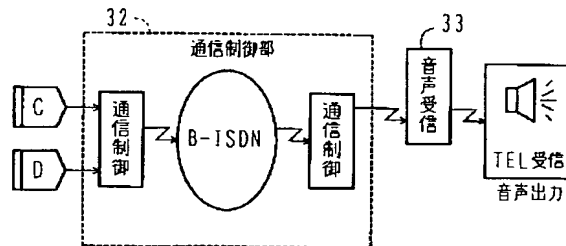
【図15】

文字入力部の処理手順



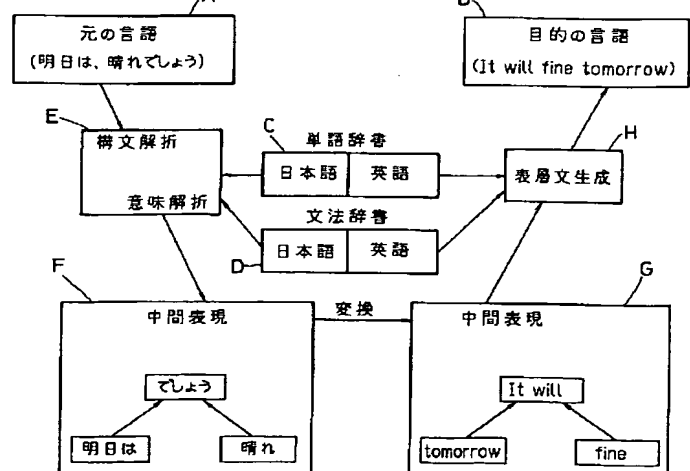
【図4】

マルチメディア公衆電話の処理ブロック図(その4)



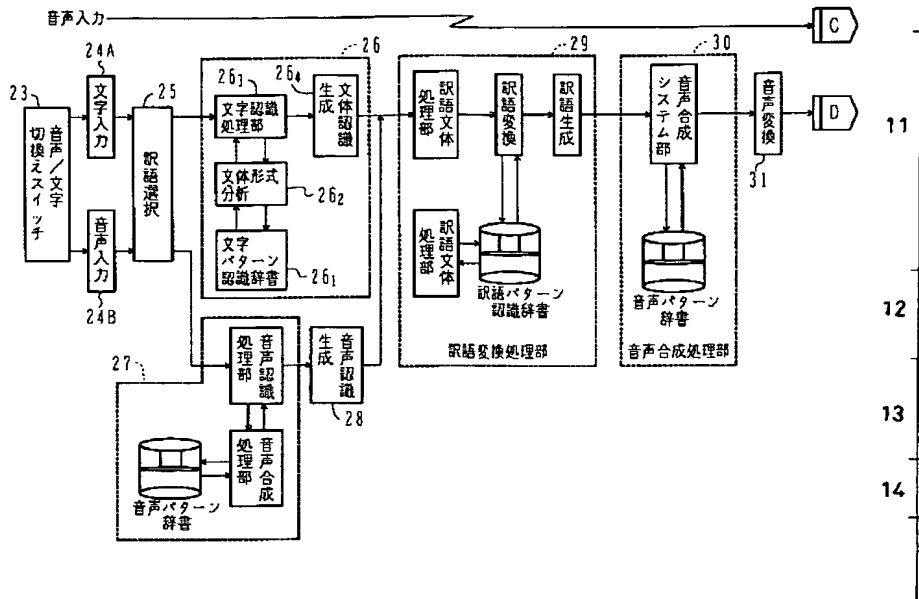
【図10】

トランスファー方式による翻訳システム構成図



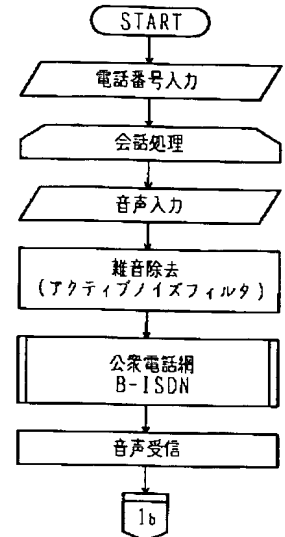
【図3】

マルチメディア公衆電話の処理ブロック図(その3)



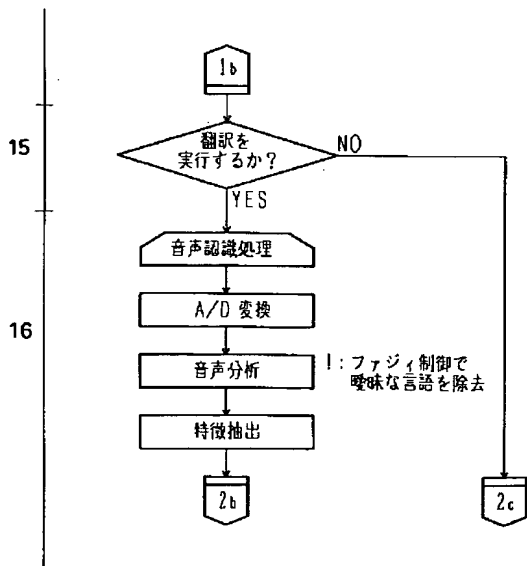
【図5】

処理フロー(その1)



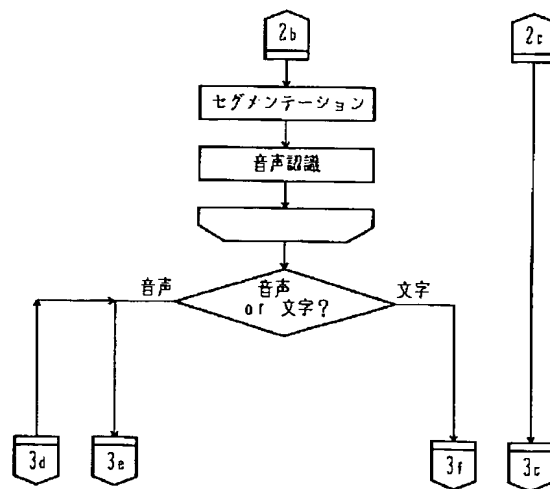
【図6】

処理フロー(その2)



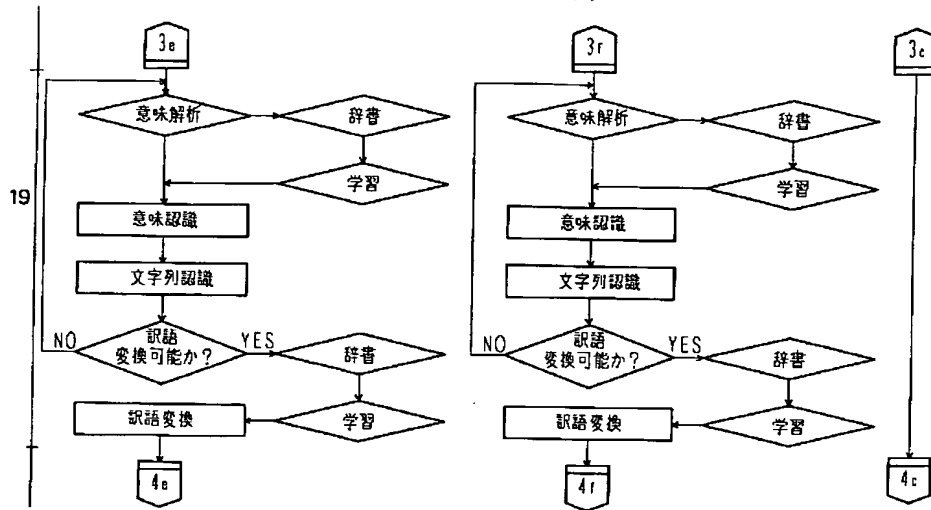
【図7】

処理フロー(その3)



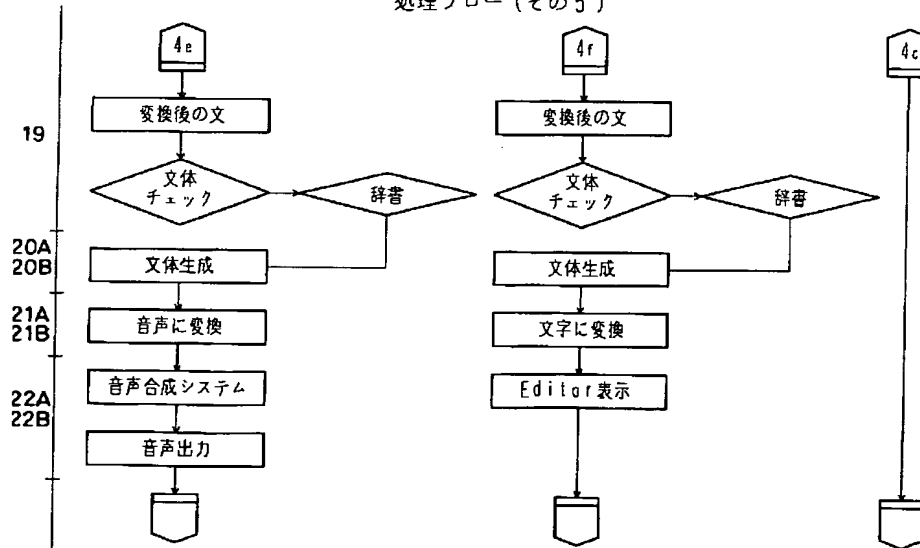
【図8】

処理フロー（その4）



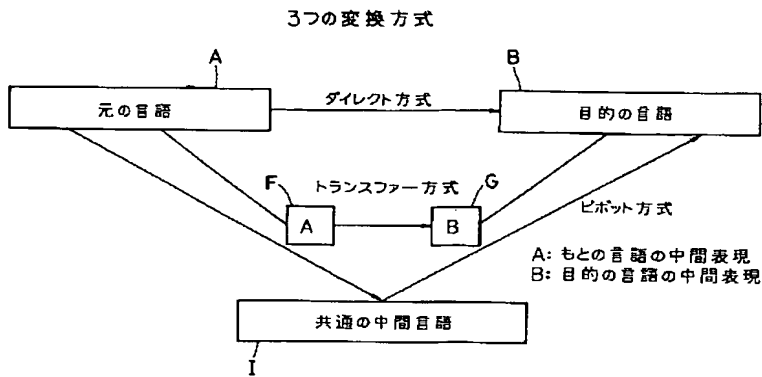
【図9】

処理フロー（その5）

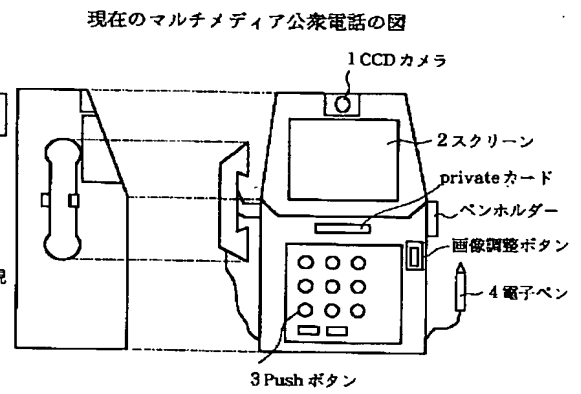




【図11】

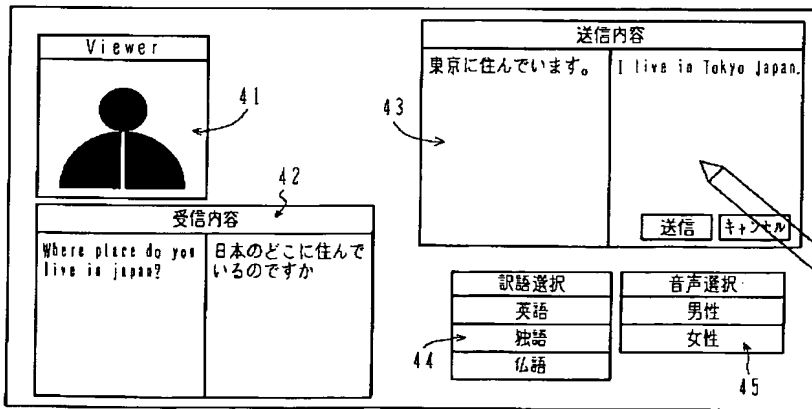


【図16】

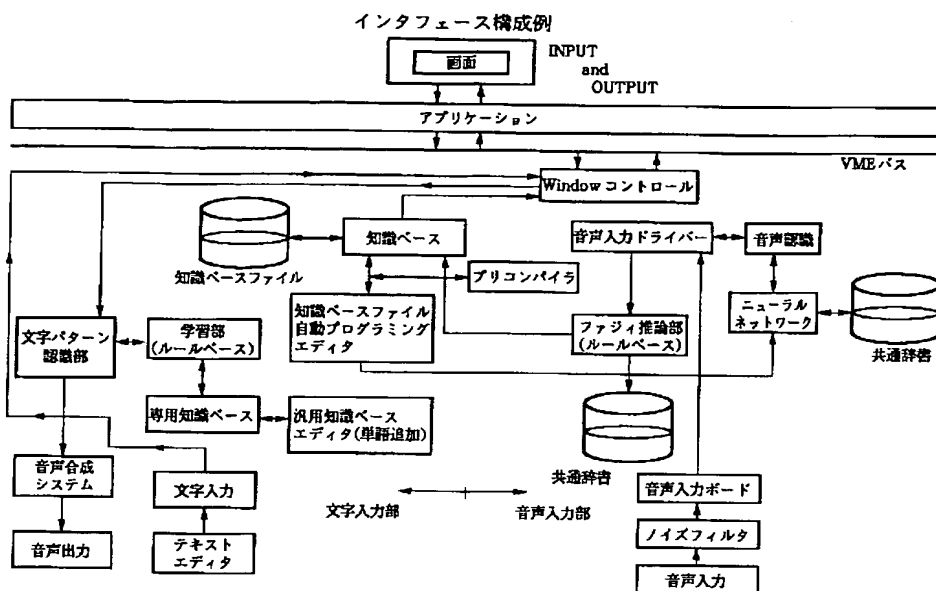


【図12】

画面レイアウト図



【例 13】



【例 14】

## 音声入力部の処理手順

